SAR Computing Accomplishments for ESS Round II: Computational Styles for Earth Science in Evolution

David W. Curkendall
Jet Propulsion laboratory
4800 Oak Grove Drive, M/S 126-104
Pasadena, CA 91109-8099
dwc@jpl.nasa.gov

***Abstract* - A basic review of the computational activities for the Round II investigation, "Advanced Computing Technology Applications to SAR Interferometry and Imaging Science" is presented - the computational techniques used and the science results obtained. The wet and dry season SAR mosaic of the Amazon will be presented as an evolution of technology for very large datasets. The computational strategies used will be contrasted with those being adopted for the Round III investigation "Cornerstone Technologies for the National Virtual Observatory". The new strategies as they can apply to SAR processing and to the access and analysis of large Earth Science datasets in general will be presented.**

I. INTRODUCTION

Since the original inception of satellite SAR technology – SeaSat back in 1978 – the reduction of the radar returns and their focusing into images has been a major preoccupation of SAR technologists and a major cost driver in the marketing and execution of any SAR mission. The lengthy time delays between the data receipt and final image and interferometric products has limited SAR's application in many cases needing near real time response – e.g. Earthquake damage assessment.

The Round II CAN was structured to explore the use of large supercomputers set in general purpose, institutionally managed contexts to attack both the issues of cost control and timely response. The 'general purpose' part replaces the historic practice of constructing a dedicated SAR processing facility for each project and seeks to participate in a cost amortization spread over many users and timescales exceeding that of a single mission. 'Large' was thought necessary to provide burst processing power that could keep up with a SAR datastream and achieve near real time results when appropriate. A conceptual fillip was added by noting the emergence of very high speed networks enabled spreading a period of intense SAR processing over several institutional computational assets, thereby not burdening any one with too great a load.

Our proposed activities included the establishment of three distinct SAR science teams, each investigating a potentially rich and rewarding use of this phenomenology. We explored:

- Monitoring snow and ice properties in alpine regions - led by UCSB. (Ref. 1-4)
- Southern California strain buildup and earthquake displacements - led by the UCSD's Scripps Institution of Oceanography. (Ref. 5-8).
- Assessment of the state and evolution of the Amazon Rain Forest. Basin - led by JPL (Ref. 9-13).

II. RESULTS

I. Computational

The problem of the original focus was modeled after the SIR-C flights back in the 1994 time periods. The conventionally dedicated processing facility, built around the then current CM-5 featured what was determined to be a two gigaflop processor. This facility was capable of processing a complete SIR-C scene in just under 3/4 of an hour (42 min.). Since there were more than 15,000 of these scenes, the complete processing of a ten day SIR-C mission was measured in years ~ 1330 eight hour shifts! The Round II goals of achieving 100 Gigaflops could conceptually reduce this time by a factor of 50 to 216 hours – i.e. with around–the–clock processing, you could reduce the data to images in somewhat less than the original flight duration. Under these circumstances, one could imagine a steady state 'free flyer' and produce all the products in near real–time.

During the three year duration of the project, the scientific emphasis and goals within the Earth Enterprise changed to the point where it no longer looked reasonable to focus on the multi-channel, multi-polarization SIR-C like mission as a free flyer. Instead, the increasing success of Repeat Pass Interferometry, InSAR, (the European ERS missions led the way here) made the effective processing of interferometry data type much more relevant. Accordingly, after the 50 Gigaflop milestone was met, the SAR Computational team retired the scalable SIR-C processor and instead focused on the parallelization and speed-up of JPL's Repeat Orbit Interferometry software, or ROI PAC.

In early May of 1999, we achieved the 100 Gigaflop milestone with the fully parallelized ROI PAC software that carried the computations past the original image formation phase (image formation completed the SIR-C computational chain) and through to the correlation of image pairs and final interferogram formation. The scaling chart is shown in Fig. 1

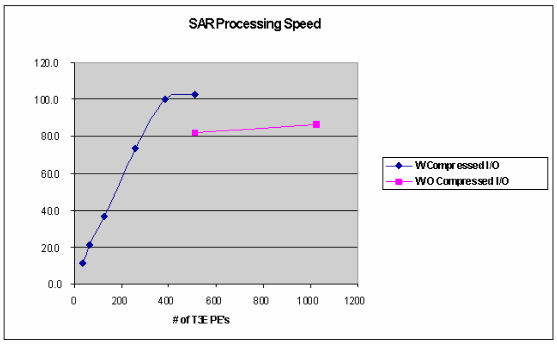

Fig. 1. 100 Gigaflop Scaling Chart

The ultimate performance achieved was limited more by the I/O characteristics of the T3E than by any computational constraints. The effective goal in processing of the raw data was to completely process some 14.6 mbytes/s of input data through to the interferogram output. This is a small fraction of the actual machine I/O needed as the original 4 bit data samples are expanded into double precision calculations and many intermediate products must be temporarily saved as the computation proceeds. All told in order to meet the 100 gigaflop goal with the first complete scalable ROI-PAC package, we needed to obtain an average disk I/O performance in the aggregate of 181 megabytes/s; the best we achieved was an I/O rate of about 156 mbytes/s [1024 processors, pink curve]. In order to meet the 100 gigaflop goal, we added data compression to the ROI-PAC calculation stream for all intermediate I/O steps and achieved a total I/O requirement reduction of about 40% [at the expense, of course, of additional calculations]. This was sufficient to speed up the overall processing so that the 100 GFOP goal – 50 times the original SIR-C processor rate – was achieved.

Although no experiments could be run, there was little doubt that with a more capable I/O system, we could have demonstrated 100 times the SIR-C processor rate. There is much detail here, but the Round II SAR investigation has conclusively shown that it is possible to perform the very complex computations needed for both SAR and InSAR at potential rates exceeding those implied by keeping up with a continuous data stream.

The experiments to demonstrate that this processing rate could be spread over several disparate computational complexes was less successful. Figure 2 illustrates a highly desirable end product capability that features data receipt from a 5 meter antenna at Scripps and the networked ability to distribute the data over storage facilities and multiple computers shown here, e.g., at Caltech, JPL and GSFC. Each of the elements of this system was demonstrated individually:

- Receipt of data and realtime movement to JPL for processing
- Portability of the scalable ROI PAC and operation on Cray, HP, and SGI parallel platforms.
- Image and interferogram results visualization at the JPL Powerwall facility.

But the Round II investigation was completed before the envisioned system was demonstrated as an integrated whole suitable for both high intensity (process all of the mission's data) and interactive (process selected frames with repeating pass pairs selected from the realtime downlink and from archival storage) processing.

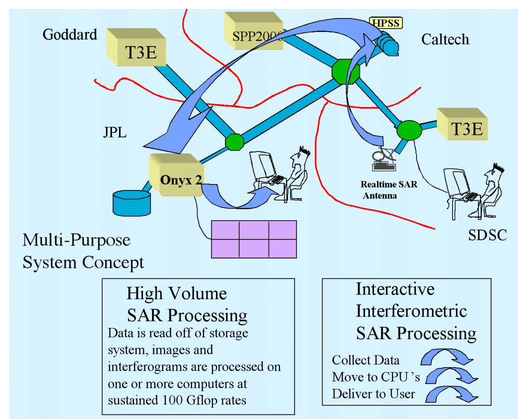

Fig. 2 Distributing the SAR Computations Geographically

II. Science

All three Science sub-teams produced excellent results. The details are beyond the scope of the present paper, but the reader is encouraged to visit http://alphabits.jpl.nasa.gov/SAR/final.html and peruse the Final Sub-Team Reports and the references cited here.

Computationally, the Amazon Science Project, Arvores, necessitated the development of both scalable mosaic creation and subsequent visualization softwares. For the first time, a continental scale dual season mosaic was created at the 100m resolution level from some 2,220 individual Japanese Earth Resources Satellite. The geodetic control of each of these mosaics was held to the sub-pixel level so that both the visualization and the subsequent science interpretations of inundation and seasonal change could be seen and computed on a pixel by pixel basis. This mosaic is being distributed widely and is, itself, the subject of several papers.

A very tiny version (to conform to this paper's limitations of 5 mbyte total) of the entire dual season mosaic is shown below: The original composite image is over 15 Gbytes in size and encompasses the entire Amazon rainforest. In the Figure the dry season maps to blue and the wet season is directed equally to red and green. Greyscale then indicates no seasonal or temporal change; the various subtle colors encode various changing phenomena (the bright ones are differences in coverage).

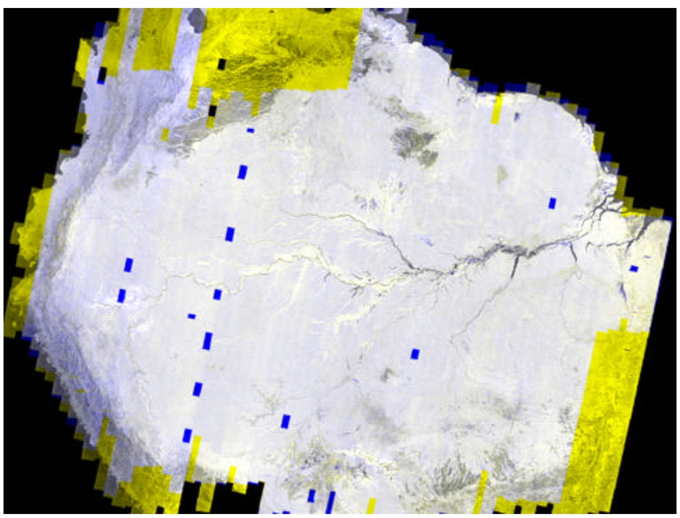

Fig 3. Dual Season Amazon Mosaic

III. Processing for Current and Future Missions – SRTM & Echo

A. SRTM

The Shuttle RADAR Topography Mission (SRTM) is a follow-on to the SIR-C missions but uniquely employs single pass SAR interferometry to obtain a topographic map of the entire world up to the latitudes reached by the Shuttle orbit. It is a joint NASA NIMA mission and because of security reasons had to configure a dedicated processor facility within a single secure complex. There are significant elements to the geodetic data processing beyond the interferogram formation stage pursued in Round II and discussed here. The production processing for this mission was to have commenced in mid September of last year. But after the 9/11 event, NIMA made several requests for special regional processing and the full scale production start date was delayed until mid April '02. It is expected to take approximately one year to complete.

B. Echo

It is hoped that NASA will soon mount a free flying InSAR satellite. A current JPL led proposal under consideration, named Echo, is planned as a multi-year InSAR ongoing mission. It is positioned as an Earth change and hazard observatory. As such it will focus on the detection and monitoring of changes in the solid Earth and ice and monitoring potential hazards such as seismic and volcanic.

Echo as a project will not be responsible for the data processing and production of the science products. Instead Project Echo will focus on gathering the data and distributing it to multiple mirrored archives for easy access and long term active archiving. Versions of ROI PAC will be freely distributed and individual science efforts can be initiated by members of the Science Team and by the broader scientific community.

Potential Echo results could be enhanced significantly if a way could be found to implement a high intensity and interactive distributed processing facility similar to the depiction in Figure 2. Such networks of computing assets are now functionally enhanced by an emergent body of middleware software falling under the rubric of the Global Grid (http://www.gridforum.org/). Indeed, Ames Research Center leads NASA's participation in these developments with its own in house project called the Information Power Grid, IPG, http://www.ipg.nasa.gov/.

A Space Science Computational Technologies Round III investigation of the "High Performance Cornerstone for the National Virtual Observatory, NVO" is layering its effort on the Grid infrastructure as is a companion program mounted by the NSF. In this way, it is planned to make the vast archives of astronomical data both more scientifically accessible and more closely aligned with the large amount of computing cycles to process them. At the heart of the NVO concept is the notion of 'multi-sky–survey analysis' enabling the gathering of the multiple datasets from each of the geographically dispersed archives and bringing them to a central point of analysis. Although the NVO does not prescribe a rigid architecture, it is expected that the most common modality of access and computational direction will be though WEB Portals, making access very democratic and uniform over the entire astronomical community.

Thus far, however, the Earth Science Enterprise has not embraced an equivalent effort for its own access and processing of Earth Science data. We would propose that a Grid based system for the Echo mission could serve as an excellent prototype to demonstrate the advantages of the emerging Grid infrastructure to dramatically improve access both to large, Earth science data sets and to the means of their analyses.

If such a system were to be deployed and made available for Echo processing, access could still be broadly based as is planned for now. But instead of leaving each scientist to devise his own local processing center, appeal could be made to the more collective Grid based assets so that all three of the main ingredients – the data, the cpu cycles, and most importantly, the software – reside on the Grid and are easily accessed.

Echo goals would be strengthened; the amount of data actually processed would increase dramatically, even though it would remain individually directed. It might further be possible to reduce processing duplication if it could also be arranged to make individually computed products deposited back into the archive and made available.

The types of studies that could be mounted would similarly broaden, making, e.g., the seismic (and aseismic) monitoring of very wide areas on the scale of the tectonic plates themselves possible. In emergencies, a large Earthquake perhaps, rapid and thorough analysis would be enhanced; the engines of computation would be there and could scalably be devoted to the problem in the amounts needed. The data would be secure because of Echo's inherently distributed and redundant depository design. But additionally, even if the Earthquake damaged some Grid assets, the undamaged portions would still provide the necessary facilities for rapid response.

It would not be difficult; all the building blocks are essentially in place. What is needed is the effort and the will to piece them together and bring such a system into being.

ACKNOWLEDGMENT

The author of this essentially review article wishes to acknowledge the very rewarding experience it was to work with the SAR Grand Challenge Team. All members contributed to the products discussed here in various ways. Of particular importance to the development of some of the themes of this paper were Herb Siegel, Craig Miller, Paul Siqueria & Bruce Chapman from JPL, Dave Sandwell and Bernard Minster from Scripps and J. Shi and Jeff Dozier from UCSB. We wish also to thank the GSFC/Cray Computational Support Team for there inestimable help in successfully meeting the performance milestones.

REFERENCES

[1] J. Shi, S. Hensley, and J. Dozier, "Mapping snow cover with repeat pass synthetic aperture radar", Proceedings IGARSS '97, IEEE No. 97CH36042, vol. II, pp. 628-630, 1997.

[2] J. Shi and J. Dozier, "Estimation of Snow Water Equivalence Using SIR- C/X-SAR, Part I: Inferring snow density and subsurface properties", in press, IEEE Transactions on Geoscience and Remote Sensing, 1999.

[3] J. Shi and J. Dozier, "Estimation of Snow Water Equivalence Using SIR- C/X-SAR, Part II: Inferring snow depth and particle size", in press, IEEE Transactions on Geoscience and Remote Sensing, 1999.

[4] J. Shi and J. van Zyl, "On improvement of bare surface soil moisture estimation", Proceedings IGARSS '98, IEEE No. 98CH36174, vol. IV, pp. 1834- 1836, 1998.

[5] Price, E. J. and D. T. Sandwell, Small-scale deformation associated with the Landers 1992 California earthquake mapped by InSAR Phase Gradient, J. Geophys. Res., v.103, p. 27001-27016, 1998.

[6] Sandwell, D. T. and E. J. Price, Phase gradient approach to stacking interferograms, J. Geophys. Res., v. 103, p. 30,183-30,204, 1998.

[7] Sandwell, D. T., L. Sichoix, D. Agnew, Y. Bock, and J-B. Minster, Near- real-time radar interferometry of the Mw 7.1 Hector Mine Earthquake, Geophys. Res., Lett., in press, May, 2000.

[8] Sandwell, D.T. and L. Sichoix, Topographic phase recovery from stacked ERS interferometry and a low resolution digital elevation model, J. Geophys. Res., submitted February 25, 2000.

[9] Siqueira, P., Chapman, B., Saatchi, S., and Freeman, A., "Amazon Rainforest visualization/Classification Enabled by Supercomputers (ARVORES)", 1997, IGARSS97, Singapore.

[10] Siqueira, P., Chapman, B., McDonald, K., and McGarragh, G., "Post-processing Calibration and Evaluation of SAR radiometry", 2000, IGARSS2000, Honolulu, Hawaii.

[11] Siqueira, P. Hensley, S., Shaffer, S., Hess, L., McGarragh, G., Chapman, B., Holt, J., and Freeman, A., "A continental Scale mosaic of the Amazon Basin using JERS-1 SAR", accepted for publication, IEEE Transactions Geoscience and Remote Sensing, 2000.

[12] Anthony Freeman, Bruce Chapman, and Paul Siqueira, "The JERS-1 Amazon Multi-Season Mapping Study (JAMMS): Science Objectives and Implications for Future Missions", submitted to the International Journal of Remote Sensing, 2000.

[13] Bruce Chapman, Paul Siqueira, and Anthony Freeman, "The JERS Amazon multi-season mapping study (JAMMS): Observation Strategies and Data Characteristics", submitted to the International Journal of Remote Sensing, 2000.